

# Database of Covid-19 cases and deaths (and tests) by age and sex

Tim Riffe, Enrique Acosta

20 April, 2020

# Status

- ▶ No name. CAgeDB?
- ▶ Shared publicly here:  
[https://www.github.com/timriffe/covid\\_age](https://www.github.com/timriffe/covid_age)
- ▶ In continuous development

# Collecting

*all measures cumulative*

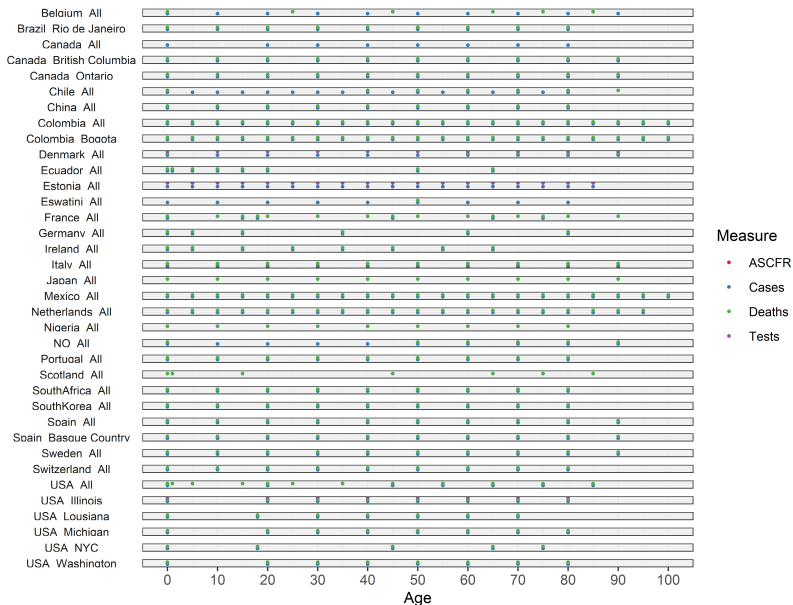
- ▶ Case counts by age and sex
- ▶ Covid-19 death counts by age (and sex if avail)
- ▶ Test counts by age (and sex if avail)

(all from health ministries)

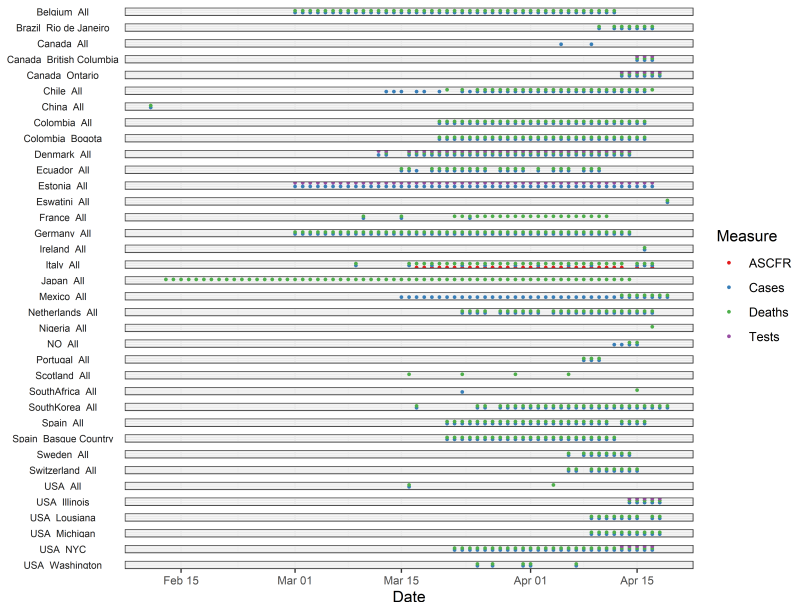
# Coverage

- ▶ 5 Continents
- ▶ 27 Countries
- ▶ 11 subnational populations

# Coverage



# Coverage



# DB Structure

- ▶ Manual entry in source-specific Drive templates and databases (decentralized)
- ▶ Entry databases collected in R
- ▶ Processing in R (code on [github](#))
- ▶ Provisional `inputDB` and output in 5 and 10 year age groups

# Input format

```
> head(inputDB)
```

```
# A tibble: 6 x 11
```

	Country	Region	Code	Date	Sex	Age	AgeInt	Metric	Measure	Value
	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	<dbl>
1	Belgium	All	BE01.03.2020	01.03.2020	b	0	10	Count	Cases	0
2	Belgium	All	BE01.03.2020	01.03.2020	b	10	10	Count	Cases	4
3	Belgium	All	BE01.03.2020	01.03.2020	b	20	10	Count	Cases	1
4	Belgium	All	BE01.03.2020	01.03.2020	b	30	10	Count	Cases	1
5	Belgium	All	BE01.03.2020	01.03.2020	b	40	10	Count	Cases	5
6	Belgium	All	BE01.03.2020	01.03.2020	b	50	10	Count	Cases	3



# Output format

```
> head(DB)
```

```
# A tibble: 6 x 9
```

	Country	Code	Date	Sex	Age	AgeInt	Cases	Deaths	Tests	
	<chr>	<chr>	<chr>	<chr>	<int>	<dbl>	<dbl>	<dbl>	<dbl>	
1	Belgium	BE	05.04.2020	05.04.2020	b	0	10	136.	0.192	NA
2	Belgium	BE	05.04.2020	05.04.2020	b	10	10	202.	0.420	NA
3	Belgium	BE	05.04.2020	05.04.2020	b	20	10	<u>1</u> 725.	1.11	NA
4	Belgium	BE	05.04.2020	05.04.2020	b	30	10	<u>2</u> 279.	3.32	NA
5	Belgium	BE	05.04.2020	05.04.2020	b	40	10	<u>3</u> 119.	13.0	NA
6	Belgium	BE	05.04.2020	05.04.2020	b	50	10	<u>3</u> 767.	51.9	NA

# The team

- ▶ 35+ collaborators

## Special roles and teams

- ▶ Coordination
- ▶ Methods design and validation
- ▶ Automation
- ▶ Source compilation and search
- ▶ Quality control
- ▶ Trainer
- ▶ Population structures

# Value added

- ▶ Central source
- ▶ Harmonize measures to counts
- ▶ Harmonize age groups
- ▶ Reproducible
- ▶ Transparent, inclusive, dynamic

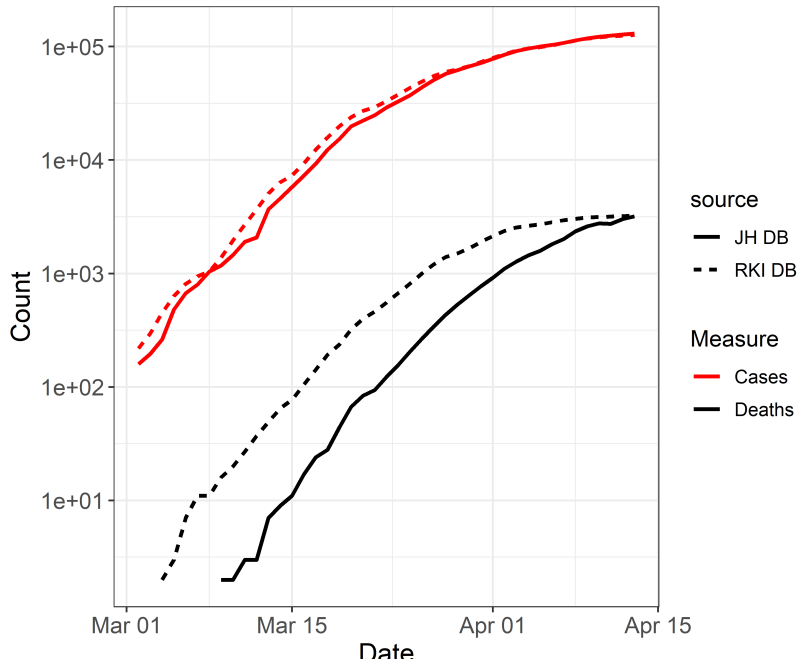
# Limitations

- ▶ Inconsistent definitions
- ▶ Inconsistent reporting
- ▶ Late registration and other lags
- ▶ Testing differences

# What we can do

- ▶ Renewing sources (e.g. Germany, Washington)
- ▶ Standard documentation
- ▶ Technical reports
- ▶ Extra adjusted series

## Renewing sources example



# What users can do

- ▶ Model biases
- ▶ Merge with other databases
- ▶ Use indirect methods, be creative